

# PROYECTO FINAL CURSO DATA ANALYTICS



MIGUEL GARCÍA MORENO

# ÍNDICE

Abstract.....	3
Introducción .....	4
Objetivos y descripción del caso práctico.....	5
Apartado ETL.....	6
Apartado SQL .....	11
Respuestas a las preguntas SQL.....	12
Apartado visualización .....	17
Conclusiones apartado práctico .....	20
Bibliografía.....	21

## Abstract

This project is presented as a final project for the data analytics course by EOI in collaboration with the governal programme Garantía Juvenil. The end of uploading this work is to showcase the depht and transversality level adquired during this two hundred and fifty hours long course. It is worth mentioning the fact that this is an end to end poryect, in the sense that we were only provided with the website where data was hosted. Of course, there were a set of goals that needed to be met, however, what must be highlighted is that this proyect has been carried out without external aid.

This work covers several fields in data analytics as it takes on downloading data from a link in a website using ETL tools and storing this information in a database via ETL procedures. Once data has been stored properly, SQL queries are used for extracting pertinent information in order to answer some questions. Eventually, Power BI's dashboards allow us to collect insights from this massive data set and support decision making for executive teams.

This document shows the step by step process in detail using tools in their community/free version. Codes of the programs can be found alongside this document in my linkedIn profile.

## Introducción

Este proyecto se presenta como trabajo final del curso de data analytics impartido por la EOI en colaboración con el programa de Garantía Juvenil. La motivación para subir este proyecto es mostrar el nivel de profundidad y transversalidad con la hemos trabajado en este curso. Es digno de mención el hecho de que este trabajo es de tipo end to end, en el sentido que lo único proporcionado por el profesorado fue la página web desde dónde se deben descargar los datos. Por supuesto, también existían unas pautas con puntos que debían ser satisfechos, pero lo destacable es que el proyecto se lleva a cabo de principio a fin por el propio alumnado sin intervención externa. Es, por tanto, un trabajo exhaustivo pues toca todas las etapas de procesamiento de datos; comenzando por la extracción, continuando con el almacenamiento, y finalizando con la lectura de los datos en la base de datos junto con la elaboración de dashboards que apoyen la toma de decisiones.

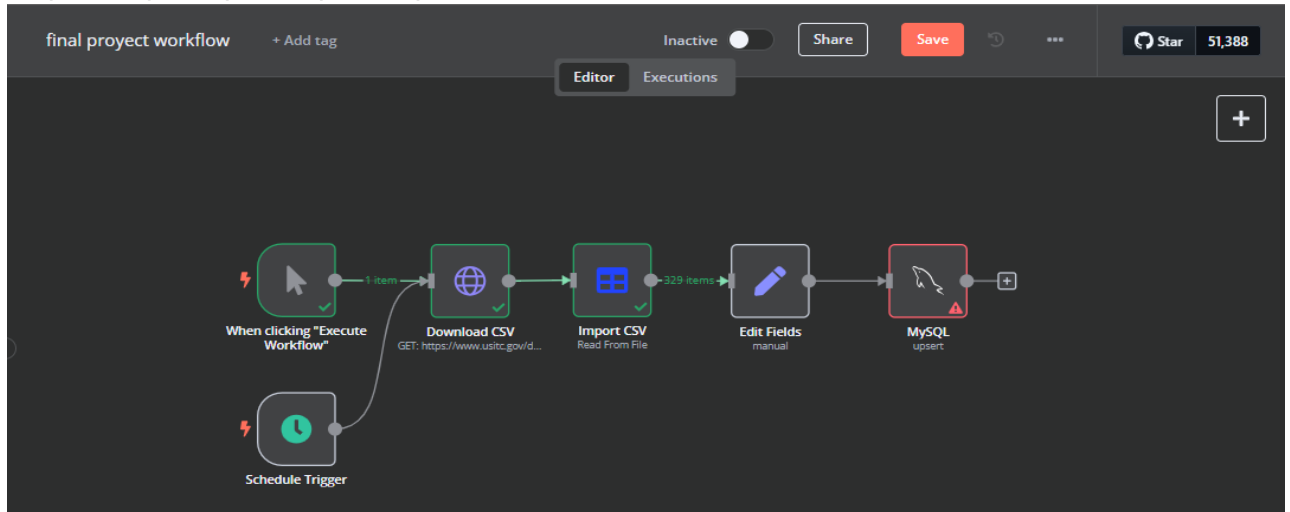
## Objetivos y descripción del caso práctico

Los objetivos de este proyecto son la aplicación transversal de técnicas inherentes al análisis de datos desde cero hasta la elaboración de un dashboard que permita facilitar la toma de decisiones del personal ejecutivo en la materia. Más concretamente, a partir de un enlace web se extrae el conjunto de datos por procedimientos de ETL. Igualmente, a través de ETL se cargan los datos a una base de datos, MySQL. Una vez almacenada esta información, se realizan queries en SQL para responder a cuestiones relativas al conjunto de datos. Por último, se crea un dashboard con Power BI para extraer de forma adecuada toda la información que un conjunto masivo de datos ofrece y exponerla en un modo fácilmente comprensible e intuitivo. Pudiendo así, servir como una base sólida sobre la que apoyarse en la toma de decisiones.

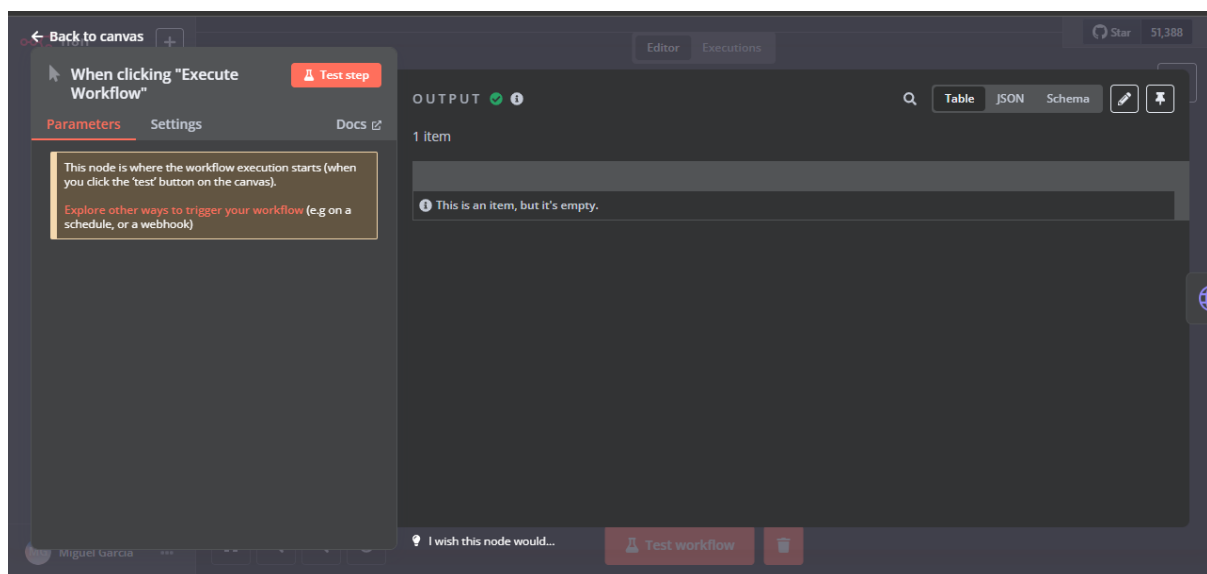
Los datos versan sobre el beneficio económico, empleo e inversiones multinacionales de empresas con inversiones directas en el extranjero. Estas empresas comprenden actividades económicas en 185 países, 25 industrias y durante un periodo de tiempo de 12 años. Entre los datos encontramos columnas relativas a los nombres y abreviaturas de los países de origen y destino. Encontramos identificadores naics de los sectores junto con sus descripciones. Así como también encontramos columnas que hacen referencia a la totalidad de afiliaciones, a las de tipo greenfield (construcción de las operaciones de la empresa en el extranjero desde el principio, es decir, crear una empresa completamente nueva en el país destino) y a las de tipo merger (cualquier afiliación que se produzca como resultado de una adquisición o fusión de las empresas). Podemos visualizar también las inversiones en las columnas de assets (activos en inglés) y el número de empleados según el tipo de afiliación. Añadir parte de objetivos y mejorar descripción

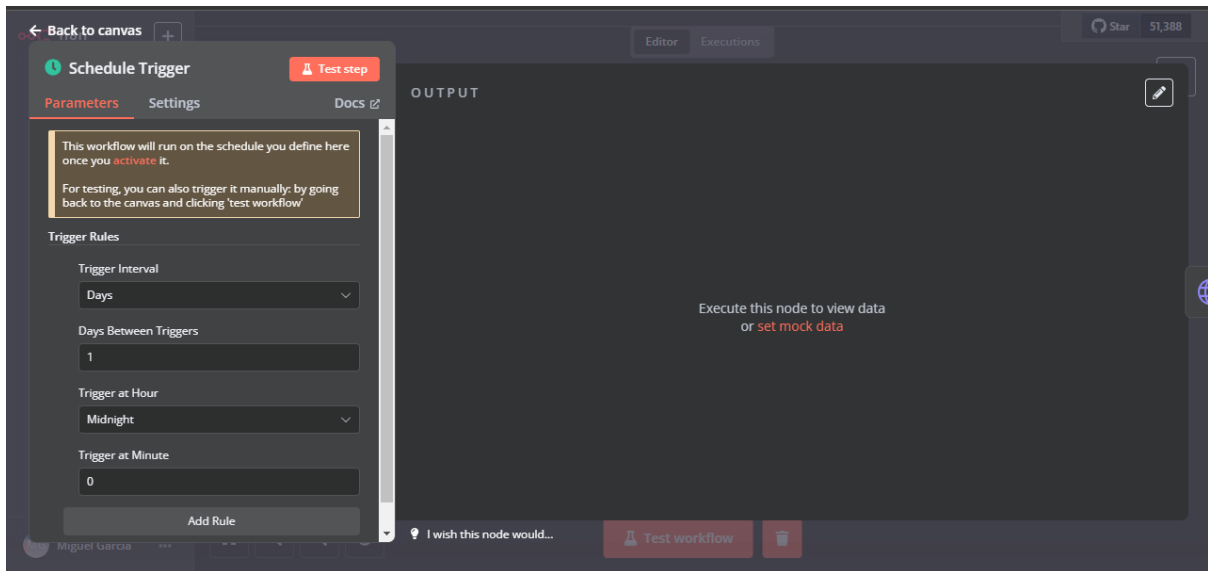
## Apartado ETL

Se pretende por medio de la herramienta n8n automatizar la recolección de los datos desde la página web dada por el ejercicio a la base de datos de MySQL. Con tal de ser lo más fiel a la realidad y pensando en una implementación general he construido el siguiente esquema que explicaré paso a paso.



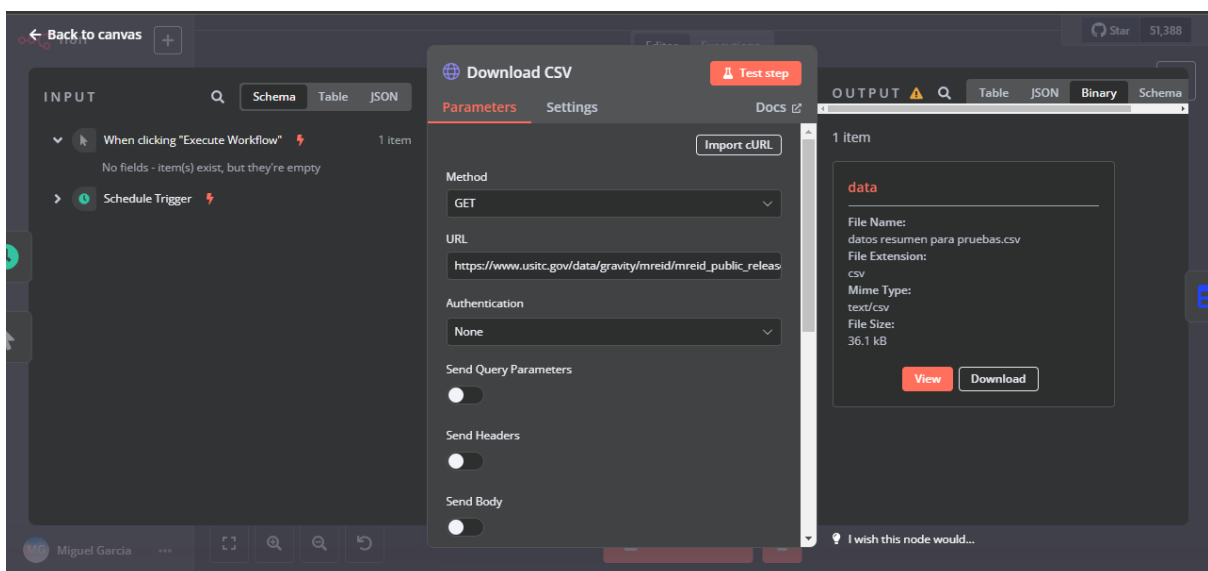
Como puede observar el lector, hay dos triggers, el primero de ellos automático y el segundo manual. En la segunda etapa encontramos la instrucción que nos permitirá descargar el fichero csv de la página web. Continuando por el work flow, leemos ese fichero csv previamente descargado e introducimos una etapa de manipulación de esos datos, en la cual me extiendo más adelante. Por último, cargamos los datos a la base de MySQL. Primero veremos los triggers





El trigger manual no tiene ninguna especificación, en cuanto a la automática la he establecido en una activación diaria a la medianoche para que inicie el proceso de actualización de datos. Evidentemente esto es puramente simbólico pues en concreto esta página web no continúa actualizando estos datos.

En la segunda etapa tenemos



Se puede ver que el enlace es en efecto el que nos ocupa. Por otro lado, el fichero en output es uno de prueba que sirvió para realizar ajustes y comprobar el correcto funcionamiento del work flow.

A continuación, leemos el fichero csv

iso3_o	country_o	iso3_d	country_d
ABW	Aruba	NLD	Netherlands
ABW	Aruba	NLD	Netherlands
ABW	Aruba	NLD	Netherlands
ABW	Aruba	NLD	Netherlands
ABW	Aruba	NLD	Netherlands

Tras un pequeño análisis exploratorio previo se ha comprobado que los encabezados aparecen en la primera línea, es por eso que marcamos “header row”. Obsérvese por la salida de la orden a la derecha del pantallazo que los datos se han leído correctamente y los encabezados son los adecuados.

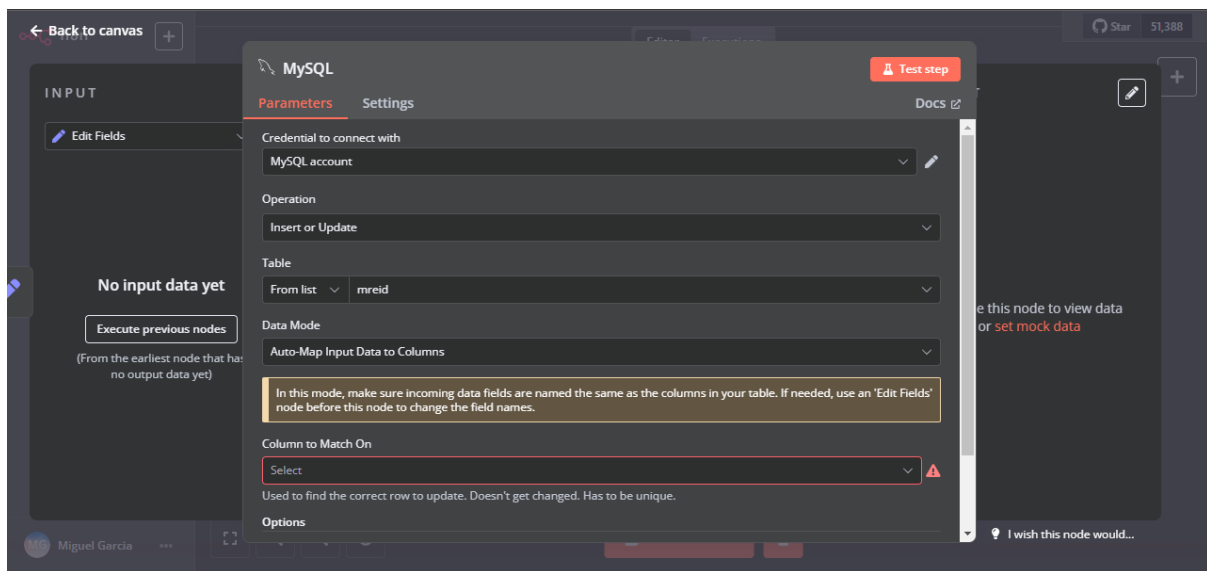
En el siguiente paso hacemos una pequeña manipulación de campos.

iso3_o	country_o	iso3_d	country_d	year
ABW	Aruba	NLD	Netherlands	201
ABW	Aruba	NLD	Netherlands	201
ABW	Aruba	NLD	Netherlands	201
ABW	Aruba	NLD	Netherlands	201
ABW	Aruba	NLD	Netherlands	201



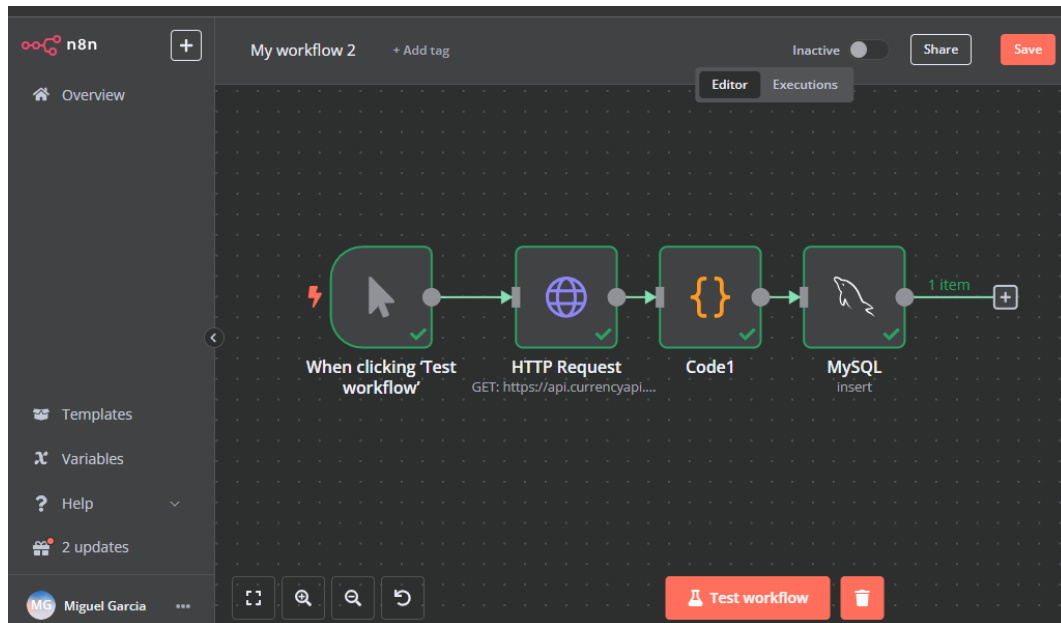
Aquí se crea un campo nuevo al que vienen en los datos, pues para actualizar la base de datos necesitamos una columna que actúe de clave primaria y permita la actualización o la inserción de nuevos casos. Sin esta columna, n8n no sabría cuáles entradas se encuentran ya en la base de datos y cuáles de ellas no, por lo que se insertarían todos los registros como nuevos, dando lugar a duplicidades y falseamiento en la obtención de conclusiones tras análisis estadísticos. Esta clave primaria la hemos formado como concatenación de cadena texto de todos los campos que no son métricas. Por supuesto en los países hemos usado sus códigos de 3 dígitos iso, en lugar de su nombre completo por dejar lo más corta posible la clave primaria. Además del campo nuevo se mantienen en el work flow todos los demás campos.

Por último, realizamos un upsert en SQL



Aquí puesto que la base de datos la hemos creado con los nombres de columna exactamente coincidentes con los que se dan los datos en el csv podemos marcar en “data mode” la opción “auto map” pues no tendrá inconvenientes en detectar a qué columna de la base de datos va cada columna del csv. En este caso para “column to match on” debería seleccionarse la columna PrimaryKey en la base de datos. Así por lo ya expuesto en el anterior párrafo, los datos se actualizarían o ingresarían, según si fueran nuevos o no, identificados por su clave primaria. En el pantallazo no marco la columna de la base de datos pues esta columna no existe, ya que por las propias limitaciones de n8n nos hemos visto forzados a cargar de forma manual los datos directamente en MySQL.

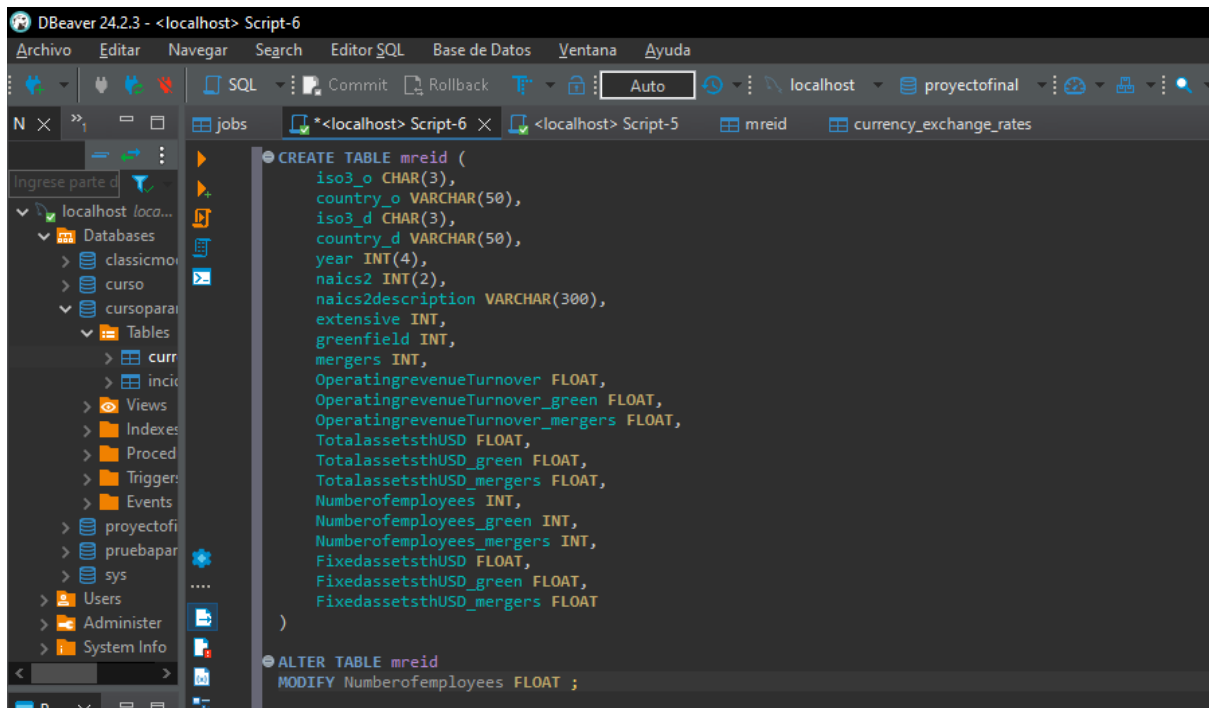
Para concluir el apartado de ETL se mostrará el diagrama usado para extracción a través de APIs de los tipos de cambio entre las divisas euro y dollar.



Pormenorizando el esquema, se aprecia un trigger manual que desencadena el work flow. Este trigger acciona una petición de los datos de tipos de cambio mediante API. Seguidamente existe una etapa de código para modificar el archivo json recibido de la orden anterior y adecuarlo a las características de la tabla de MySQL que va a recibir esta información. En última instancia se carga este archivo a la base de datos.

## Apartado SQL

Aquí trataremos en primer lugar la creación de la tabla para el proyecto con lenguaje del tipo DDL en MySQL. Con ayuda de la documentación del conjunto de datos podemos saber qué tipo de variable necesita ser declarada para cada una de las columnas. En este caso no ha hecho falta declarar columnas con la propiedad de clave primaria o secundaria puesto que no la hay. De la misma forma no ha hecho falta la introducción de restricciones más allá de la propia tipología del dato.



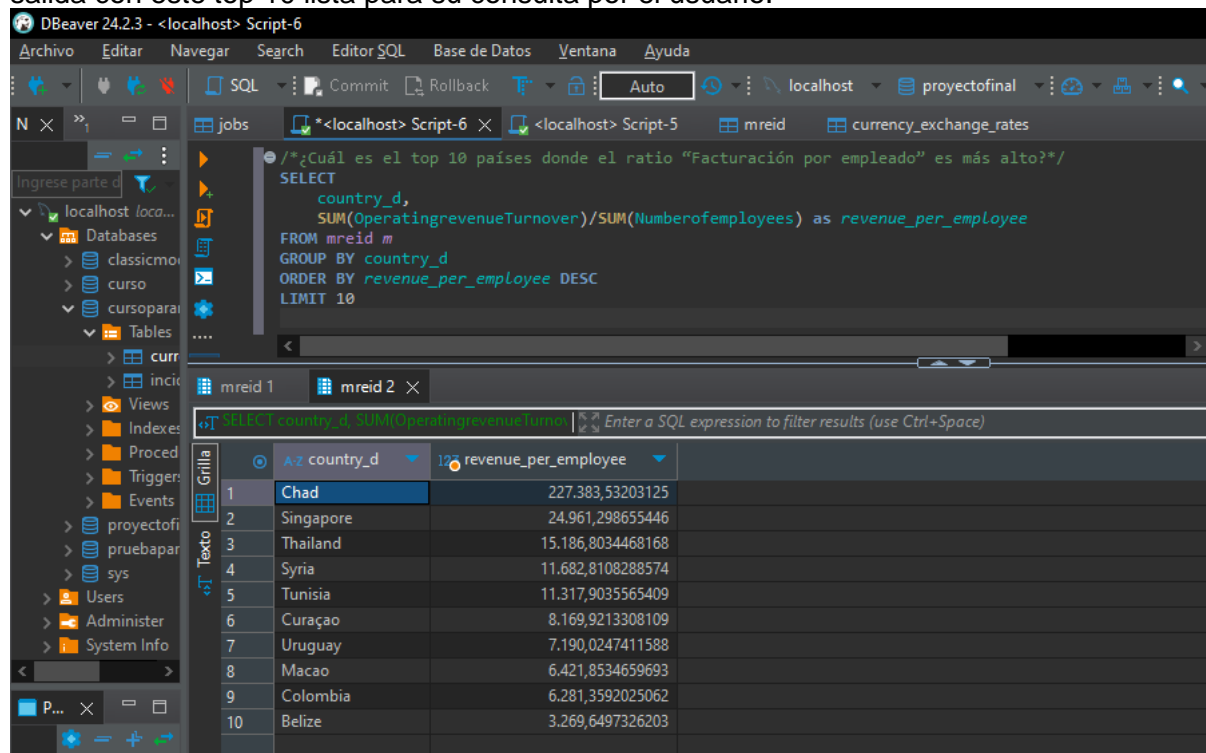
En la captura se muestra la creación de dicha tabla, a la que se la denomina mreid. A posteriori tuvo que modificarse la columna Numberofemployees porque a pesar de indicar en naturaleza que debe tratarse de una variable de tipo INT, el conjunto de datos contenía valores de tipo FLOAT en esa columna.

Una vez creada la tabla se han cargado los datos a MySQL y se ha procedido a responder las preguntas a través de queries.

## Respuestas a las preguntas SQL

1. ¿Cuál es el top 10 países donde el ratio “Facturación por empleado” es más alto?

En la query seleccionamos la columna `countries_d` pues es en los países destino en los que están trabajando estos empleados y dividimos la suma de facturación (`operatingrevenueTurnover`) entre la suma de empleados (`numberofemployees`). Además, a la columna de facturación por empleado le proporcionamos un nombre representativo e imponemos un criterio de ordenación para establecer el ranking. Obsérvese que aparece la salida con este top 10 lista para su consulta por el usuario.



The screenshot shows the DBeaver 24.2.3 interface. The SQL editor contains the following query:

```
/*¿Cuál es el top 10 países donde el ratio “Facturación por empleado” es más alto?*/
SELECT
  country_d,
  SUM(OperatingrevenueTurnover)/SUM(Numberofemployees) as revenue_per_employee
FROM mreid m
GROUP BY country_d
ORDER BY revenue_per_employee DESC
LIMIT 10
```

The results are displayed in a table with the following data:

	country_d	revenue_per_employee
1	Chad	227.383,53203125
2	Singapore	24.961,298655446
3	Thailand	15.186,8034468168
4	Syria	11.682,8108288574
5	Tunisia	11.317,9035565409
6	Curaçao	8.169,9213308109
7	Uruguay	7.190,0247411588
8	Macao	6.421,8534659693
9	Colombia	6.281,3592025062
10	Belize	3.269,6497326203

2. ¿Cuál es el bottom 10 países donde el ratio “Facturación por empleado” es más bajo?

En este apartado, repetimos como en el anterior, pero ordenando en sentido inverso. Al encontrarnos valores null procedentes de divisiones por 0, cambiamos los valores null por el valor numérico 0. Y de nuevo obtenemos un ranking, que en términos reales nos da una lista de países sin empleados físicos.

DBEaver 24.2.3 - <localhost> Script-6

Archivo Editar Navegar Search Editor SQL Base de Datos Ventana Ayuda

SQL Commit Rollback Auto localhost proyectofinal

/\*¿Cuál es el bottom 10 países donde el ratio "Facturación por empleado" es más bajo?\*/

```
SELECT
  country_d,
  IFNULL(SUM(OperatingrevenueTurnover)/SUM(Numeroofemployees),0) as revenue_per_employee
FROM mreid m
GROUP BY country_d
ORDER BY revenue_per_employee
LIMIT 10
```

mreid 1 mreid 2 X

SELECT country\_d, IFNULL(SUM(OperatingrevenueTurnover)/SUM(Numeroofemployees),0) as revenue\_per\_employee

	A2 country_d	12 revenue_per_employee
1	Maldives	0
2	Saint Lucia	0
3	Nicaragua	0
4	South Sudan	0
5	Seychelles	0
6	Andorra	0
7	Honduras	0
8	Lesotho	0
9	Anguilla	0
10	Gibraltar	0

3. ¿Cuáles son los 5 sectores con mejor evolución entre 2010 y 2021, en el número de empresas (extensive) en inversión (TotalassetsthUSD) doméstica en España?

En esta pregunta se pretende mostrar como resultado una tabla con los datos necesarios para representar un gráfico de líneas tal como se pide en las especificaciones del proyecto. Por esto mismo no se muestran en ranking y agrupados únicamente por año por sector. Se puede ver en la captura de pantalla mostrada un renombramiento de la columna `naics2description` para un reconocimiento más veloz, así como una cláusula `where` detallando los criterios de filtrado. Como son inversiones domésticas se exige que el país de origen y el de destino sean el mismo y que sean forzosamente España. Por otro lado, para el total de la inversión de las empresas extensive se ha usado `TotalassetsthUSD` y también se ha mostrado la columna agregada de extensive.

localhost> Script-6

Navegar Search Editor SQL Base de Datos Ventana Ayuda

SQL Commit Rollback Auto localhost proyectofinal <localhost> consultas\_classicmodels

Proyectos

/\* ¿Cuáles son los 5 sectores con mejor evolución entre 2010 y 2021, en el número de empresas (extensive) en inversión (TotalassetsthUSD) doméstica en España? \*/

```
SELECT
  naics2description as sector ,
  year,
  SUM(TotalassetsthUSD) as investments_total,
  SUM(extensive) as number_of_enterprises
FROM mreid m
WHERE (country_o = country_d) AND (country_o = 'Spain')
GROUP BY year, naics2description
```

mreid 1 mreid 2

Enter a SQL expression to filter results (use Ctrl+Space)

	sector	year	investments_total	number_of_enterprises
1	Agriculture, Forestry, Fishing and Hunting	2.010	358.059,5625	578
2	Oil and Gas Extraction	2.010	1.495,745,75	162
3	Utilities	2.010	8.255.298,5	1.081
4	Construction of Buildings	2.010	3.970.074,25	1.591
5	Food and Textile Manufacturing	2.010	2.786.016,5	847
6	Materials Manufacturing	2.010	4.041.477,75	766
7	Finished Product Manufacturing	2.010	2.421.689	1.042
8	Wholesale Trade	2.010	3.602.255	2.708
9	Food and Beverage Stores	2.010	334.648,53125	573
10	Miscellaneous Store Retailers	2.010	36.510,19921875	152
11	Air Transportation	2.010	868.284,375	913
12	Postal Service	2.010	121.144.7734375	97

Refresh Save Cancel Exportar datos ... 200 200+

200 row(s) fetched - 0,489s, on 2024-12-17 at 20:25:48

CET es Writable Smart Insert 66:34:1849 Sel: 0 | 0

4. ¿Cuáles son los 5 sectores con peor evolución entre 2010 y 2021, en el número de empresas (extensive) en inversión (TotalassetsthUSD) internacional en España? (empresas con origen internacional y con destino España)

Este apartado se resuelve de forma paralela al anterior y la tabla mostrada tiene también el mismo fin, poder dibujar un gráfico de líneas. En este caso modificamos la cláusula where pues se pide la inversión internacional. De esta forma seleccionamos por países origen aquellos que no son España y por país destino España. Por último, obsérvese que la métrica se mantiene igual pues continúan pidiendo la inversión de todas las afiliaciones.

localhost> Script-6

Navegar Search Editor SQL Base de Datos Ventana Ayuda

SQL Commit Rollback Auto localhost proyectofinal <localhost> consultas\_classicmodels

/\*¿Cuáles son los 5 sectores con peor evolución entre 2010 y 2021, en el número de empresas (extensive) en inversión (TotalassetsthUSD) internacional en España? (empresas con origen internacional y con destino España)\*/

```

SELECT
  naics2description as sector ,
  year,
  SUM(TotalassetsthUSD) as investments_total,
  SUM(extensive) as number_of_enterprises
FROM mreid m
WHERE (country_o != 'Spain') AND (country_d = 'Spain')
GROUP BY naics2description, year

```

mreid 1 mreid 2

Grilla

1 2 sector 123 year 124 investments\_total 125 number\_of\_enterprises

	sector	year	investments_total	number_of_enterprises
1	Food and Beverage Stores	2.010	658.438,59375	225
2	Finance and Insurance	2.010	30.456.801,00793457	222
3	Real Estate	2.010	6.386.783,392696381	384
4	Legal Services	2.010	13.768.228,99673748	959
5	Management of Companies and Enterprises	2.010	4.220.636,410888672	273
6	Arts, Entertainment, and Recreation	2.010	183.072,593963623	185
7	Food and Beverage Stores	2.011	4.646.994,633453369	233
8	Finance and Insurance	2.011	86.709.720,33331084	231
9	Real Estate	2.011	20.036.406,682617188	408
10	Legal Services	2.011	36.524.187,81493378	1.034
11	Management of Companies and Enterprises	2.011	65.951.453,71036267	298
12	Arts, Entertainment, and Recreation	2.011	1.641.070,318359375	200

Refresh Save Cancel Exportar datos ... 200 200+

200 row(s) fetched - 2s (0,121s fetch), on 2024-12-17 at 20:24:08

CET es Writable Smart Insert 80 : 33 : 2255 Sel: 0 | 0

5. Importe en € (EUROS), en el 2020 de la inversión extranjera en la Península Ibérica (España, Portugal y Andorra), en el sector “Finanzas y Seguros”.

Para la resolución de esta pregunta es necesario implementar una subselect, pues requiere aunar la información de sendas tablas, `currency_exchange_rates` (cuyos datos provienen de la ETL) y `mreid`. De la primera tabla interesa tomar el tipo de cambio correspondiente a 2020 y por ello se filtran los resultados por la fecha, de manera que esta sea igual a '2020-01-01 00:00:00'. En la tabla `mreid` seleccionamos la columna del nombre del sector y calculamos la suma de los valores de las inversiones. Además, filtramos con `where` que los países destino sean Portugal y España, el año 2020 y el sector Finanzas y Seguros (en inglés). Se concluye mostrando el nombre del sector y una columna calculada que aplica el tipo de cambio de USD a EUR.

The screenshot shows the DBeaver 24.2.3 interface. The left sidebar displays the database structure, including tables like 'currency\_exchange\_rates'. The main editor shows a SQL script with a comment: '/\* 5. Importe en € (EUROS), en el 2020 de la inversión extranjera en la Península Ibérica (España, Portugal y Andorra), en el sector "Finanzas y Seguros". \*/'. The script uses a subquery to select 'sector' and 'quantity' from 'mreid' and joins it with 'currency\_exchange\_rates' to calculate the value in EUR for the year 2020. The results grid at the bottom shows one row for the 'Finance and Insurance' sector with a value of 344,451,061,000,169,04.

```

GROUP BY naics2description, year

/* 5. Importe en € (EUROS), en el 2020 de la inversión extranjera en la Península Ibérica
(España, Portugal y Andorra), en el sector "Finanzas y Seguros".
*/

SELECT
  A.sector,
  A.quantity * B.EUR as quantity_in_eur
FROM
  (SELECT
    naics2description as sector,
    SUM(TotalassetsthUSD) as quantity
  FROM mreid m
  WHERE (country_o != country_d)
  AND ((country_d = 'Spain') OR (country_d = 'Portugal') OR (country_d='Andorra')))
  AND (year=2020)
  AND (naics2description = 'Finance and Insurance ')) AS A,
  (SELECT EUR
   FROM currency_exchange_rates cer
   WHERE date = '2020-01-01 00:00:00') AS B
  
```

A-z sector	123 quantity_in_eur	Valor
1 Finance and Insurance	344,451,061,000,169,04	Finance and

6. Importe en € (EUROS), en el 2015 del ratio “Facturación por empleado” en Italia, para el sector de “Bienes Raíces”.

Esta consulta vuelve a ser muy parecida a la anterior, sin embargo, cambiamos el filtro del sector al pedido en esta ocasión, así como el año. Por otro lado, se actualiza la métrica calculada para que sea Facturación por empleado como en los apartados primero y segundo.

The screenshot shows the DBeaver 24.2.3 interface with a new SQL script. The comment is: '/\*Importe en € (EUROS), en el 2015 del ratio "Facturación por empleado" en Italia, para el sector de "Bienes Raíces"\*/'. The script is similar to the previous one but filters for the year 2015 and the 'Real Estate' sector. The results grid shows one row for the 'Real Estate' sector with a value of 38,712,389,334,734,336.

```

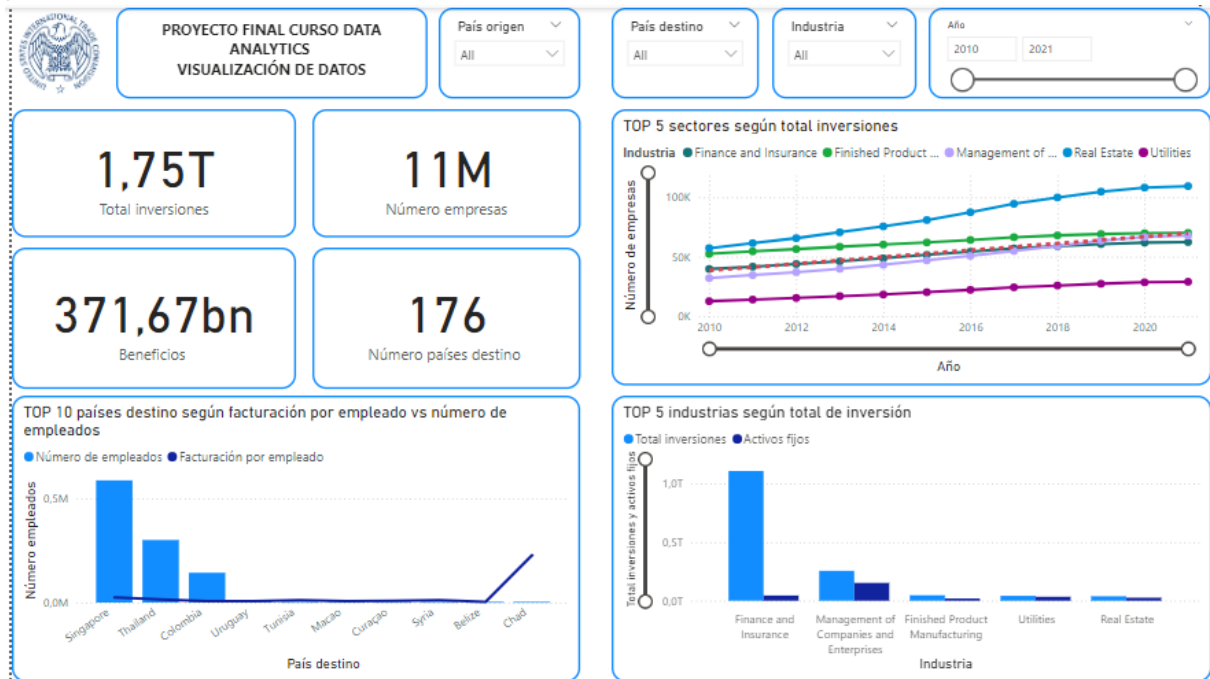
/*Importe en € (EUROS), en el 2015 del ratio "Facturación por empleado" en Italia,
para el sector de "Bienes Raíces"*/
SELECT
  A.sector,
  A.quantity * B.EUR as quantity_in_eur
FROM
  (SELECT
    naics2description as sector,
    SUM(TotalassetsthUSD) as quantity
  FROM mreid m
  WHERE (country_o != country_d)
  AND ((country_d = 'Spain') OR (country_d = 'Portugal') OR (country_d='Andorra')))
  AND (year=2015)
  AND (naics2description = 'Real Estate ')) AS A,
  (SELECT EUR
   FROM currency_exchange_rates cer
   WHERE date = '2015-01-01 00:00:00') AS B
  
```

A-z sector	123 quantity_in_eur	Valor
1 Real Estate	38,712,389,334,734,336	Real Estat



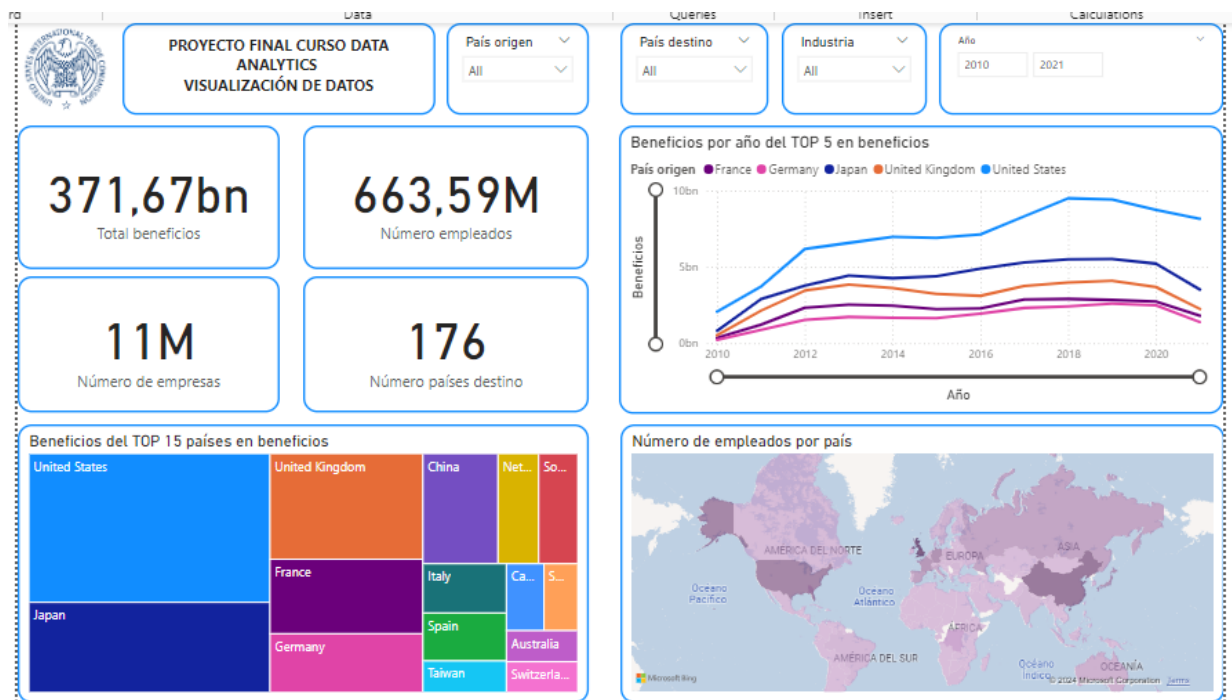
## Apartado visualización

Para la visualización de los datos en paneles de Power BI me he decidido por crear tres paneles. Veamos cada uno de ellos.



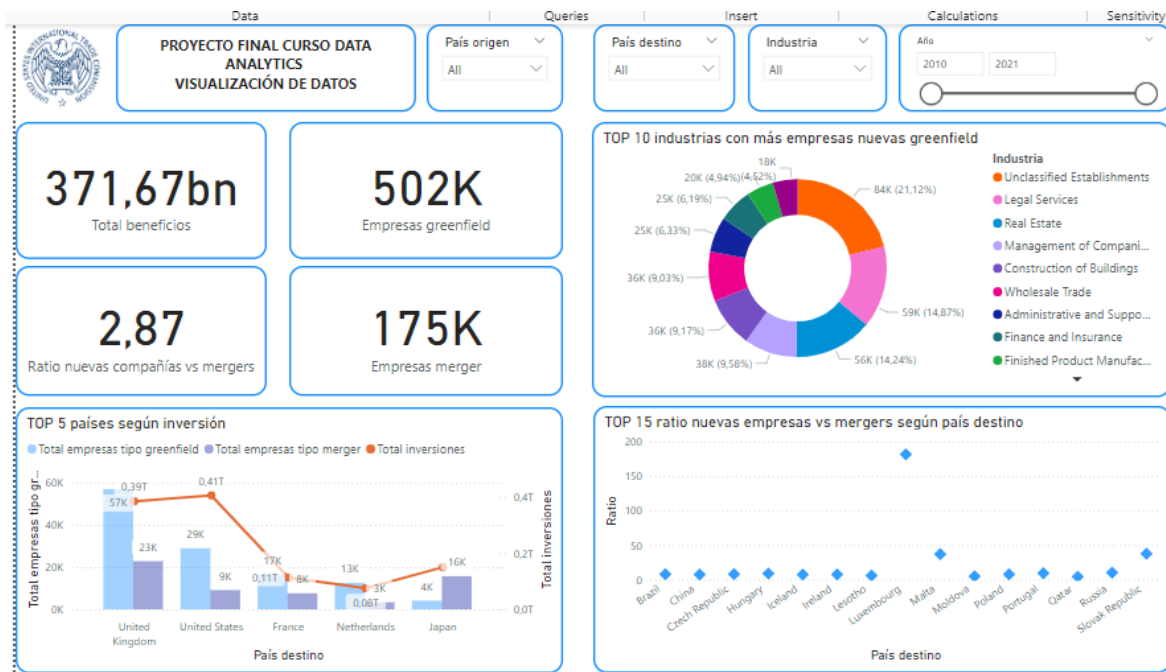
En el encabezado del panel encontramos el logotipo de la institución de la cual hemos obtenido los datos acompañados por el título del proyecto. Junto a estos dos hay cuatro paneles que nos permiten el filtrado de los datos a nuestro gusto. Este primer panel permite por sí mismo responder a las preguntas. En primer lugar, la respuesta a las primeras dos preguntas vendría recogida en “TOP 10 países según facturación por empleado vs número de empleados”. La línea azul oscura muestra el valor de facturación por empleado y además las barras azules nos informa sobre la magnitud de la fuerza de trabajo en cada uno de estos países. Continuando con la tercera y cuarta pregunta, pueden consultarse en este panel en el gráfico de líneas arriba a la derecha, que nos indica el número de empresas según sector y aparece ya filtrado a las 5 industrias con mayor inversión. Además, con ayuda del tooltip se puede visualizar la cantidad total de inversiones. Por otro lado, para seleccionar inversión doméstica o extranjera se puede hacer uso de las segmentaciones del encabezado del panel. Como observación, en este gráfico que estamos comentando se incluye una línea de tendencia que aproxima matemáticamente como es el comportamiento general de estas industrias en el tiempo. Por último, las últimas dos preguntas pueden responderse con ayuda del gráfico en la esquina izquierda y manipulando los filtros a conveniencia. En añadidura a estos tres gráficos hay distintas tarjetas que nos proporcionan información relevante sobre los datos como puede ser el total de beneficios o el número de empresas. Cabe también puntualizar que para la facturación por empleado ha sido necesaria la creación de una métrica adecuada.

Veamos ahora el segundo panel



Este segundo panel mantiene el encabezado del anterior, por lo tanto, también los filtros. Las tarjetas, sin embargo, reflejan otra información para completar. Aquí visualizamos tres gráficas interesantes en cuanto a los beneficios de las empresas. Comenzando por la esquina derecha podemos ver una gráfica de líneas en el que se representa la tendencia de las cinco economías más fuertes a lo largo del tiempo. Los márgenes móviles facilitan al usuario el análisis concreto en áreas de su elección. Por otro lado, encontramos el gráfico de la esquina izquierda abajo un Treemap que explica cómo se distribuyen los beneficios de las quince economías más potentes en beneficios. Por último, se muestra un mapa con un relleno gradiente en función del número de empleados que podemos encontrar en cada uno de estos países. De este modo rápidamente podemos localizar aquellas zonas en la que hay un mayor volumen de fuerza trabajadora. Podría sorprender tal vez que Reino Unido tenga un volumen de trabajadores similar al de China o Estados Unidos por la diferencia de población, y gracias al gradiente descubrimos esta “anomalía” visualmente de forma inmediata.

Pasemos al último panel



Este último panel se centra en la tipología de la empresa en los países destino. Podemos ver el total de beneficios, las empresas de tipo greenfield, las de tipo merger y una métrica que indica la razón de nuevas empresas sobre mergers. Así como antes, el encabezado permanece inmutable posibilitando el filtrado a elección del usuario. En primer lugar, en la esquina superior derecha hay un gráfico de sectores circular mostrando el top 10 de industrias con más empresas nuevas. En segundo lugar, en la esquina inferior izquierda hay una gráfica combinada de barras y línea que describe el número de empresas nuevas y las de tipo merger (procedentes de fusiones o adquisiciones), así como dibuja el total de inversiones con la línea naranja. Finalmente, en la esquina inferior derecha vemos un gráfico de puntos donde se pueden ver los ratios de empresas nuevas vs mergers de los 15 países destino en los que esta métrica es más elevada.

## Conclusiones apartado práctico

De las visualizaciones se pueden desprender varias conclusiones. Una de ellas es que las industrias que lideran en cuanto al total de empresas y el total de inversiones son finanzas y seguros, real estate, servicios, control de compañías y productos terminados para el consumidor. Además, por mucho es el sector de finanzas y seguros el que más empresas ocupa. De igual forma podemos ver que Chad tiene la mejor facturación por empleados pese a tener 50 empleados. Observamos también que los países más beneficiados son EEUU, Japón, Reino Unido y otros, así como que el número de empleados es bajo en el continente africano y asiático salvo por China y Japón. Ahondando más en los beneficios, al verlos contrapuestos con el tiempo hay una tendencia a la caída desde 2020, seguramente a consecuencia del covid-19. Por otro lado, los países que más invierten son Reino Unido y EEUU seguidos por Francia. De los gráficos se desprende que los sectores que más empresas nuevas crean son establecimientos sin clasificar, de servicios legales y bienes y servicios. Y, para terminar, el país en el que más empresas nuevas se crean en proporción con las fusiones y adquisición es Luxemburgo, donde por cada merger se crean 182 empresas nuevas.

## Bibliografía

Página web para la extracción de datos <https://www.usitc.gov/data/gravity/mreid.htm>  
Manual de MySQL <https://dev.mysql.com/doc/refman/8.4/en/introduction.html>  
Manual de Power BI <https://learn.microsoft.com/en-us/power-bi/>